Estimating prevalence of distant metastatic breast cancer: a means of filling a data gap

Authors: Mark S. Clements, David M. Roder, Xue Qin Yu, Sam Egger, Dianne L. O'Connell.

Mark S. Clements.

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australia.

Email: mark.clements@ki.se

Telephone: +46- 8-5248 2023

Fax: +46-8-314975

David M. Roder.

National Breast and Ovarian Cancer Centre, Sydney, Australia.

Xue Qin Yu, Sam Egger.

Cancer Council NSW, Sydney, Australia.

Dianne L. O'Connell.

Cancer Council NSW, Sydney, Australia

University of Sydney, Sydney, Australia

University of New South Wales, Sydney, Australia

University of Newcastle, Newcastle, Australia.

Abstract

Purpose: To develop and validate a method for estimating numbers of people with distant cancer metastases, for evidence-based service planning.

Methods: Estimates were made employing an illness-death model with distant metastatic cancer as the illness state and site-specific mortality as an outcome, using MIAMOD software. To demonstrate the method, we estimated numbers of females alive in Australia following detection of distant metastatic breast cancer during 1980-2004, using data on patient survival from an Australian population-based cancer registry. We validated these estimates by comparing them with direct prevalence counts.

Results: Relative survival at 10 years following detection of distant metastases was low (5-20 percent), with better survival experienced by: (1) females where distant metastatic disease was detected at initial diagnosis rather than subsequently (e.g., at recurrence); (2) those diagnosed in more recent calendar years; and (3) younger age groups. For Australian females aged less than 85 years, the modeled cumulative risk of detection of distant metastatic breast cancer (either at initial diagnosis or subsequently) declined over time, but numbers of cases with this history rose, from 71 per 100,000 in 1980 to 84 per 100,000 in 2004. The model indicated that there were approximately 3 to 4 prevalent distant metastatic breast cancer cases for every breast cancer death. Comparison of estimates with direct prevalence counts showed a reasonable level of agreement.

Conclusions: The method is straightforward to apply and we recommend its use for breast and other cancers when registry data are insufficient for direct prevalence counts. This will

provide estimates of numbers of people who would need ongoing medical surveillance and

care following detection of distant metastases.

Keywords: Prevalence; Epidemiology; Metastatic cancer; Breast cancer; Statistical models.

Abbreviations: New South Wales (NSW); Mortality and Incidence Analysis Model

(MIAMOD); hazard ratio (HR); confidence interval (CI).

Purpose

The burden of cancer at a population level is often described using incidence, mortality and survival data (1). In addition, more recently, cancer registries have used prevalence of diagnosed cases as a further measure of population burden (2).

These measures are complementary in that cancer incidence would be a relevant indicator for planning services for periods immediately following diagnosis, whereas mortality data would be a relevant indicator for planning end-of-life services. By comparison, the prevalence of people surviving a cancer diagnosis at various times post diagnosis would be a relevant indicator for planning ongoing medical surveillance and treatment services throughout the disease course. In addition, prevalence data would be useful for estimating disability adjusted life years in burden-of- disease studies (3) and when estimating health costs (4).

Approaches to prevalence measurement are well established and include: (i) survey counts which can be suitable for very common cancers; (ii) cancer registry counts which is the preferred method when cancer registries are long-established and have good follow-up data; and (iii) mathematical modeling in which prevalence estimates are derived from rates of incidence and survival, which can be used when neither survey counts nor registry counts can reliably estimate prevalence (5-7).

Distant metastatic cancers, including metastases found at diagnosis and those found later in the disease course, generally would represent the most severe end of the disease spectrum in prognostic terms and a time in the disease course when resource utilization can be intensive. Partial information is available on prevalence of distant metastatic cancer from those

registries which record stage at initial diagnosis (8, 9), but prevalence of distant metastatic cancer occurring following later disease progression of localized or regional cancer is poorly recorded in most registries, if it is recorded at all.

In the absence of direct empirical evidence, means of estimating prevalence are needed, but conventional counts would rarely be applicable for distant metastatic cancer. Specifically, survey estimation generally would be problematical, due to small numbers, since many cancers do not progress to distant metastatic disease and for those that do, the median survival following distant metastatic spread would often be short. In addition, registry counts would require complete recording of disease progression through the course of the disease, which is rarely undertaken.

The method selected for prevalence estimation in this study circumvents these difficulties. It was undertaken in response to requests from consumer organizations and the National Breast and Ovarian Cancer Centre in Australia for data on the numbers of women alive following detection of distant metastatic breast cancer in Australia. The purpose was to gain evidence for service planning and advocacy and to validate the mathematical modeling by comparison with direct prevalence counts. Patient survival data were available for this study from New South Wales (NSW) which is the most populous state of Australia, including about 33% of the national population.

NSW Central Cancer Registry data indicate that most female breast cancer deaths in NSW occur in cases originally diagnosed at a localized or regional stage. For instance in 2004-2008, 4662 female breast cancer deaths were recorded, of which only 14 percent (n=662) were recorded as having distant metastases at diagnosis, while the remainder were classified

either as localized (*n*=1301, 28%), regional (*n*=1997, 43%) or of unknown stage (*n*=702, 15%) (http://tinyurl.com/3tm45we, downloaded 10/06/2012).

Our prevalence estimation methods involved reconstructing estimates of incidence and prevalence from data on cancer mortality and relative survival following detection of distant metastases. The approach did not require complete data on cancer episodes through the cancer course.

In this paper, we describe the statistical model that was used, outline our approach to estimation of model parameters, demonstrate the application of the method to estimate the prevalence of distant metastatic breast cancer for females in NSW and Australia, and validate the model by comparison of prevalence estimates with direct prevalence counts. We believe this approach would have general application in other countries and for other types of cancer where cancer registry data are insufficient for direct prevalence counts.

Methods

Model formulation

Our approach was to use the illness-death model to estimate prevalence (6). Our model specifies the following disease states: people who are "healthy or with localized or regional cancer"; an illness state which we have labeled "initial or subsequent distant metastatic cancer"; and the death state which encompasses cancer specific mortality and mortality from other causes. We use breast cancer to illustrate the methods. The transitions between states include: either a diagnosis of distant metastatic cancer from a healthy state (which we refer here to as an *initial* distant metastatic diagnosis) or the diagnosis of progression from localized or regional cancer to distant metastatic cancer (which we refer to here as a *subsequent* distant metastatic detection), with overall rate α; death from the healthy or localized/regional cancer state, with rate μ; and death from initial or subsequent distant metastatic cancer, with rate ν.

All modeled transitions would depend on age and potentially calendar year period. The excess mortality rate ν from the illness state to death would further depend on time spent in the illness state. The model is irreversible for the metastatic cancer state, such that individuals with metastatic cancer can never return to the healthy state; however, we have assumed that the mortality rates for the illness states would approach those for the healthy population as time from detection of those states increases. This irreversibility assumption further implies that we cannot model the effects of second or subsequent primary distant metastatic cancers so that any excess mortality will be attributed to the first primary distant metastasis.

We have assumed that the mortality rate for individuals who are healthy or have localized or regional cancer will be similar to the rate for the total population. This assumption implies that death due to metastatic cancer is small and that there is little or no net mortality benefit or excess mortality for people who have localized/regional cancer; that is, cause-specific mortality for localized or regional cancer is assumed to occur through progression to distant metastatic cancer. With this model, we jointly estimated the transitions from healthy and localized/regional cancer states to initial and subsequent distant metastatic cancer and we did not explicitly account for the transitions from healthy to localized or regional cancer.

This simplified model allows us to apply the **M**ortality and **I**ncidence **A**nalysis **MOD**el (MIAMOD; 5, 10). Relative survival inputs to MIAMOD were estimated by combining the initial and subsequent diagnoses of distant metastatic cancer.

Estimation using MIAMOD

We used the MIAMOD model to reconstruct distant metastatic cancer incidence and prevalence using cancer mortality rates and estimates of relative survival for distant metastatic cancer in NSW and Australia. The MIAMOD method has been used to estimate prevalence where good data were available on mortality for a whole population and survival was estimated from a sub-population; common examples include availability of nation-wide recording of deaths, but with reliable reporting of survival for only a subset of the population (11, 12). The MIAMOD method is also useful for predicting prevalence for the current year by projecting from historical data. We used relative survival estimates for NSW and applied them to estimate prevalence for NSW, for validation, and then to estimate prevalence for whole of Australia. To our knowledge, the MIAMOD model has not previously been used to

estimate numbers or prevalence rates for people in whom distant metastatic cancer has been detected.

The MIAMOD model was specified in terms of an illness-death model (6). The MIAMOD model uses the number of observed site-specific cancer deaths as the outcome in a Poisson regression, expressed as functions of the transition rates between disease states. The model predicts the number of cancer deaths given an age-cohort reconstruction for cancer incidence, using the following data: (i) population data by single year of age for single calendar years; (ii) general mortality data by single year of age for single calendar years; (iii) cancer mortality data by single year of age for single calendar years; and (iv) relative survival estimates following detection of distant metastatic cancer by single years since cancer diagnosis within aggregated age groups and aggregated calendar periods. Finally, the model requires the specification of an age-period-cohort model for distant metastatic cancer incidence.

For the analysis, distant metastatic cancer incidence in NSW was modeled in MIAMOD using age-cohort models with third order polynomials for the age effect and 0, 1 or 2 degree polynomials for the cohort effect; model selection was based on the likelihood ratio test. As a validation of the MIAMOD approach, we used the counting method to calculate the number of prevalent cases for distant metastatic breast cancer. For the counting method, all cases alive at June 15, 2004 with a previous specific cancer diagnosis were counted. The prevalence estimates from MIAMOD were found to be similar to those from the counting method for more recent years, but larger than those from the counting method in earlier years.

Data sources

NSW Central Cancer Registry data were used to estimate relative survival following initial (at diagnosis) and subsequent detection of distant metastatic disease. Incidence data for women aged 18-84 years were extracted for the period 1972-2004. Data on subsequent distant metastatic disease and survival follow-up were analyzed for the period 1980-2004. We assumed a negligible rate of distant metastatic breast cancer before 18 years of age. Following NSW Central Cancer Registry coding rules, extent of disease at initial diagnosis was based on evidence from episodes received by the Registry within 120 days of the primary diagnosis; cases of subsequent distant metastatic disease were based on episode data from 121 days of the primary diagnosis.

Approval for the analysis of the cancer data was obtained from the NSW Population and Health Services Research Ethics Committee, which waived the requirement for informed consent on the condition that the researchers received only de-identified data.

Data for relative survival were prepared as per recommendations from the Australasian Association of Cancer Registries (13). The at-risk cohort for the relative survival analysis was defined as incident breast cancer diagnoses during 1972-2004 which had either an initial diagnosis of distant metastatic breast cancer or subsequent detection of distant metastatic breast cancer based on episode data during 1980-2004. Follow-up was censored at the earlier of either age 85 years or the date December 31, 2004. Events were defined as deaths due to all causes.

Estimates of expected survival based on all-causes mortality were provided by the Australian Institute of Health and Welfare. Population-based general mortality and breast cancer

mortality data were obtained from the Australian Bureau of Statistics. These data were stratified by single calendar years (1980-2004) and by single years of age (0-84 years). Data were available for NSW and for the whole of Australia.

Relative survival estimation

We calculated relative survival using the Pohar Perme actuarial estimator (14) with a cohort-based analysis (15). Descriptive results were stratified by age (18-49 years, 50-69 years, 70-84 years), and initial or subsequent distant metastatic disease. The actuarial estimates of relative survival were calculated using the relsurv library in R version 2.13.0.

A mixture cure model of relative survival (16, 17) was used: (i) to describe differences in survival by calendar period and detection of initial or subsequent distant metastatic cancer; and (ii) to provide estimated survival parameter inputs for the MIAMOD model. Relative survival $S(t)$ to time t was modeled parametrically using $S(t) = \left(\pi + (1-\pi)S_u(t)\right)^{\delta}$, where π is the fraction cured, S_u is the survival function for those not cured and δ models the overall excess hazard ratio. The function S_u was assumed to follow a Weibull distribution with shape parameter a and scale parameter b. The parameter δ for overall change depended on calendar period and initial or subsequent metastatic disease, where $\delta = \exp(\sum_j \theta_j x_j)$ for covariates indexed by j with value x_j and parameter θ_j. The mixture cure model was: (i) implemented in R for individual-level data using maximum likelihood estimation, for model selection and estimating hazard ratios; and (ii) fitted to the actuarial estimates using weighted non-linear least squares, using SAS code provided by Roberta De Angelis, for input to MIAMOD. For the individual-level data, following notation of Lambert et al (18), the log-likelihood for individual i with event indicator d_i (=1 for an event, otherwise 0) is

$$\log(L_i) = d_i \log\left(h^*(t_i) + \frac{\delta(1-\pi)f_u(t_i)}{\pi + (1-\pi)S_u(t_i)}\right) + \delta \log(\pi + (1-\pi)S_u(t_i)) + \log(S^*(t_i))$$

where $h^*(t)$ is the background mortality rate, $S^*(t)$ is the background survival function and $f_u(t)$ is the time-to-event density function for those not cured.

For the inputs into MIAMOD, we used a model with age-specific parameters for π, a and b for those aged 18-49 years, 50-69 years and 70-84 years and for a temporal trend using δ. We used the actuarial estimates of relative survival as inputs to MIAMOD in a sensitivity analysis; the model predictions were similar for both sets of estimates.

To validate the MIAMOD approach, we also used the counting method to calculate the number of prevalent cases with metastatic breast cancer in NSW. For the counting method, all cases who were alive at 15 June 2004 with a previous metastatic breast cancer diagnosis were counted. The validity of the counting method depends on both cancer diagnosis and death matching being complete.

Recipes for applying these methods

For cancer registries that are interested in applying these methods, there are several approaches available. First, if good data are available on initial and subsequent distant metastases, then either the counting method or our reconstruction using MIAMOD can be used to estimate the prevalence of distant metastatic cancer. Second, many registries have survival data available for cases who have distant metastatic cancer at initial diagnosis, but do not have data on survival following progression to subsequent distant metastatic cancer with follow-up to death. For those registries, we propose the following steps:

13

1. Estimate relative survival for cases with initial distant metastatic cancer, using either the mixture cure model or actuarial estimators;

2. Calculate the average log-hazard ratio $\log(\overline{HR})$ across initial and subsequent distant metastatic cancer by $\log(\overline{HR}) = (1 - p)\log(HR)$, where p is the proportion of cancer-specific deaths who were metastatic at initial cancer diagnosis and HR is the age-specific hazard ratio for subsequent distant metastatic cancer compared with initial distant metastatic cancer from the Results. An alternative, less robust estimate for p is the incidence rate for initial distant metastatic cancer divided by cancer-specific mortality rate;

3. Using steps (1) and (2), calculate the population-level relative survival for cases with distant metastatic cancer. For the mixture cure model, $\log(\overline{HR})$ can be included as an offset in the overall power term. For the actuarial estimators, the adjusted relative survival estimate is calculated by raising the relative survival estimate by the power of the mean hazard ratio $\overline{HR}$;

4. Estimate the incidence and prevalence of distant metastatic cancer using MIAMOD, as described in the section *Estimation using MIAMOD*.

Third, for cancer registries without survival or without staging information, the ratio of cancer deaths to prevalent distant metastatic cases by age and time since metastatic diagnosis could be taken from a similar population. We show an example of this approach in Table 3.

Results

There were 69,690 incident breast cancers (i.e. new diagnoses irrespective of cancer stage) recorded by the NSW Central Cancer Registry for women aged 18-84 years during 1980-2004. Breast cancer incidence rates increased across this period. Of these cases, 4.6 percent (n=3180) were initially diagnosed with distant metastases, whereas 11.6 percent (n=8064) had an unknown stage.

The cohort for the relative survival analysis of distant metastatic cancer included 12,571 women aged 18-84 years with a notification for distant metastatic disease during 1980-2004 (Table 1), of whom 10,136 had died. There was only moderate variability in reported initial distant metastatic diagnoses across the study period. A greater variation was observed in the reporting of subsequent distant metastatic disease detection from the mid 1980s to the late 1990s, however, which occurred at the same time as changes in NSW cancer registry policies on data entry. As a consequence, a total of 3180 cases were observed for 1980-1984, falling to 1091 and 1129 cases in 1985-1989 and 1990-1994, rising to 2709 and 4462 cases in 1995-1999 and 2004-2004, respectively (Table 1).

[Table 1 here]

Across the period 1980-2004, approximately half of the breast cancer deaths had a recorded distant metastatic episode prior to death, with 92 percent of breast cancer deaths having a distant metastatic episode recorded during the period 1980-1984, compared with a

corresponding proportion less than 50 percent during the period 1985-1999 and 70-80 percent during the period 2000-2004 (Table 1).

Relative survival for metastatic breast cancer

Initially, we stratified the groups by age group, calendar period and initial/subsequent distant metastatic disease detection. The relative survival estimates are shown in Figure 1and Table 2. Relative survival following detection of distant metastatic breast cancer was poor, with a rapid decline in the first 5 years in all groups. Ten-year relative survival was generally 10-20 percent following initial distant metastatic breast cancer diagnosis and 5-10 percent following subsequent detection of distant metastatic disease. The possible increase in relative survival for older age groups beyond 10 years from detection of metastatic disease may be explained by less effective death matching or smaller numbers of individuals with follow-up.

[Figure 1 here]

[Table 2 here]

We initially investigated whether relative survival depended on the calendar period of distant metastatic detection. Compared with the period 1980-1984, the excess hazards of dying from all causes were higher in 1985-1994, with similar hazards applying in 1995-1999 and lower hazards in 2000-2004. Given the issue of incomplete episode data, particularly for 1985-1994, we chose to only use the data for 1980-1984 and 1995-2004 for analyses of relative survival and prevalence.

We used Akaike's Information Criterion to select a model for relative survival that included age-specific parameters for the fraction cured and the survival function for those not cured,

with a common overall trend for calendar period. We also assumed that relative survival was constant before 1980.

For estimating excess hazards by age, calendar period and for initial or subsequent detection of distant metastatic cancer, we assumed a common cure fraction and survival distribution for those not cured. The excess hazards were lower for: those women with an initial distant metastatic diagnosis (excess hazard ratio (HR)=0.57; 95% confidence interval (CI): 0.53, 0.60) compared with subsequent distant metastatic detection; for distant metastatic detection in more recent calendar periods (HR=0.93 per 10 years; 95% CI: 0.91, 0.95); and for cases aged 18-49 years (HR=0.70; 95% CI: 0.66, 0.79) and 50-69 years (HR=0.79; 95% CI: 0.75, 0.83) compared with those aged 70-84 years. Finally, there was strong evidence for effect modification for the excess hazard ratio comparing subsequent with initial distant metastatic detection by age group, with increasing hazard ratios from ages 18-49 years (HR=0.47; 95% CI: 0.41, 0.54), ages 50-69 years (HR=0.53; 95% CI: 0.49, 0.78), and ages 70-84 years (HR=0.70; 95% CI: 0.63, 0.78).

MIAMOD predictions for incidence and prevalence

The MIAMOD model was fitted to breast cancer deaths for New South Wales (Table 1) and for the whole of Australia. Based on a likelihood ratio test, an age-cohort model of incidence with a third order polynomial for the age effect and a second order polynomial for the cohort effect was the best model fit from MIAMOD. This model was then used to predict distant metastatic breast cancer incidence and prevalence in NSW and Australia.

To validate the MIAMOD modeling approach, we used the simple counting method to estimate the numbers of metastatic cases that were alive in NSW in June 2004 (Figure 2). Our

a priori expectation was that the observed counts using the counting method would under-estimate the number of prevalent cases, given that the counting method relies on complete recording of stage at primary diagnosis and the episode data for subsequent development of metastatic disease. For 2004, there was reasonable agreement between the two methods for metastatic disease diagnosed in the previous year; the estimates from the MIAMOD model were larger for diagnoses between 6 years and 20 years prior; and the estimates from the MIAMOD model were smaller for diagnoses 2-6 years prior and 21-24 years prior. The differences occurring at 2-6 years prior could be due to recent improvements in survival that are not accounted for in the modeled survival or random variation between years. Similarly, we calculated the number of prevalent metastatic cases by age group for NSW using both MIAMOD and the counting method. There was good agreement between the two methods at younger ages; for older ages, MIAMOD predicted larger numbers of women living with metastatic breast cancer. The discrepancy at older ages may be due to MIAMOD taking better account of metastatic diagnoses that were more than 5 years in the past. These findings imply that: (i) the MIAMOD model provides reasonable estimates of prevalence, although it is possible that the model underestimates survival and prevalence for the most recent calendar years; (ii) the recording of initial and subsequent metastatic disease is reasonably complete in recent years.

[Figure 2 here]

For distant metastatic breast cancer incidence in Australia, the MIAMOD model estimated that there were 2902 (95% CI: 2850, 2954) new cases in women aged 18-84 years in 2004 (Table 3), an increase from 2108 (95% CI: 2074, 2142) new cases in 1980. We found evidence for a downward trend in the cumulative risk to age 85 years of incident distant

metastatic breast cancer across the period, with 41.2 per 1000 in 1980 declining to 34.9 per 1000 in 2004, giving an average decline of 6 percent per 10 years. The MIAMOD model predicted that there were 8284 (95% CI: 8196, 8372) prevalent cases of distant metastatic breast cancer in women aged 18-84 years in NSW in 2004, which was 3.4 times the expected number of breast cancer deaths for that year.

[Table 3 here]

The number of prevalent distant metastatic cases had risen sharply from 1980, where there were 5144 (95% CI: 5016, 5271) cases, which was 2.8 times the expected number of breast cancer deaths. The crude prevalence of distant metastatic disease for women aged 18-84 years rose slowly over time, from 0.71 per 1000 in 1980 to 0.84 per 1000 in 2004, which constituted a 7 percent increase per 10 years.

Of those women predicted to be living with distant metastatic breast cancer, 73 percent were less than 10 years from the time of their distant metastatic diagnosis. The proportion of women who had lived with distant metastatic breast cancer for less than 10 years declined with age, with this proportion being more than 95 percent for those aged less than 40 years, but lowered to 58 percent for those aged 80-84 years.

[Table 4 here]

If the ratio of prevalent cases to deaths and the proportion of prevalent distant metastatic cases by year since metastatic diagnosis could be generalized between populations, then we can estimate the numbers of prevalent metastatic cases by the years since metastatic diagnosis

using: the observed numbers of breast cancer deaths times the modeled ratio of prevalent

cases to breast cancer deaths times the proportion of cases by time since distant metastatic

diagnosis (Table 4). As an example, for every hundred breast cancer deaths aged 50-54 years,

we would expect approximately 214 (=3.19×67), 61, 29, 13 and 6 prevalent cases with 0-4, 5-

9, 10-14, 15-19 and 20+ years since metastatic diagnosis, respectively.

In a sensitivity analysis, we assessed whether the prevalence estimates were affected by

assuming no trend in survival prior to 1980. We re-fitted the model assuming that the trend in

survival from 1980 to 2004 continued prior to 1980, such that earlier periods had worse

survival. The revised model provided very similar estimates of prevalence for 2004 and

incidence across 1980-2004, while the model predicted lower prevalence for 1980, with 4525

prevalent cases as compared with 5144 from the main model.

Conclusions

We have applied an illness-death model to estimate the prevalence of distant metastatic cancer, with distant metastatic cancer comprising the illness state and site-specific mortality being an outcome. The method was implemented using MIAMOD software and did not require complete data on cancer staging or disease progression. Reasonable agreement was found between estimates from the MIAMOD method and direct prevalence counts.

We have provided several approaches for applying the method and we believe that they would have most general applicability in cancer registries that collect information on stage or degree of spread. The resulting prevalence estimates would address an important gap in registry data where registry data are insufficient for direct prevalence counts. The modeled estimates would indicate numbers of cancer cases in need of special surveillance and ongoing services following detection of distant metastases.

To illustrate the method, we applied the model using NSW Central Cancer Registry breast cancer survival data. We found relative survival for distant metastatic breast cancer to be lower for older women and those with subsequent distant metastatic breast cancer compared with women with distant metastatic disease detected at initial diagnosis. This survival improved over the period 1980-2004. Distant metastatic breast cancer can be aggressive with rapid fatality, such that a relative survival of only 5-20 percent applied at 10 years following detection.

Using the MIAMOD model to reconstruct distant metastatic breast cancer incidence for women in the whole of Australia aged 18-84 years, we found that the incidence had declined across the period 1980-2004, while prevalence had increased slowly. We made the potentially useful observation that there were 3 to 4 prevalent cases of distant metastatic breast cancer per breast cancer death.

There were some limitations in the analysis. First, we did not include women aged 85 years and over at diagnosis. There is considerable uncertainty around a number of critical parameters for this older age group, so we did not expect estimates for this group to be reliable. Second, we did not present results for the fraction cured. Second, although the relative survival model provided estimates of cure, the fraction surviving after 10 years was low and breast cancer is well recognized as having long-term excess mortality (19). Third, we assumed that all deaths recorded as breast cancer were due to distant metastatic breast cancer. This implies that, for localized or regional breast cancer cases, excess breast cancer deaths would only be through progression to metastatic cancer. Any other excess mortality, such as related to treatment of the cancer, would be either small in number or recorded as due to other causes of death. Fourth, the accuracy of the relative survival estimates may depend on the completeness of staging. We had good evidence for completeness of the NSW data on patient survival during 1980-1984 and 2000-2004, based on the proportion of breast cancer deaths with a recorded metastatic episode. We cannot offer strong advice on the effect of missingness patterns for staging data, but urge caution in assuming such data are missing completely at random. As an extension, the relative survival estimates could be weighted by inverse selection probabilities to better account for differential selection between initial and subsequent distant metastatic cases, with robust variance estimation using the sandwich estimator.

We provided a several approaches to estimate the prevalence of distant metastatic breast cancer for sub-populations. As a possible extension of this model application, further analyses could describe the rate of progression from localized and regional disease to metastatic disease.

The method is straightforward to apply. We recommend that cancer registries use it for estimating the prevalence of distant metastatic disease for evidence based service planning when registry data are insufficient for direct prevalence counts. We also recommend that the validity of resulting estimates be confirmed by direct counting of people with a history of distant metastatic disease, where registries have the data to undertake these counts, and that further efforts be made to refine the model.

Conflicts of interest: Professor Roder was an employee of the National Breast and Ovarian Cancer Centre.

References

1. Ries LAG. (2007) Cancer survival among adults : U.S. SEER program, 1988-2001, patient and tumor characteristics. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute.

2. Capocaccia R, Colonna M, Corazziari I, et al. (2002) Measuring cancer prevalence in Europe: the EUROPREVAL project. Ann Oncol. 13: 831-9.

3. Barendregt JJ, Van Oortmarssen GJ, Vos T, Murray CJ. (2003) A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. Popul Health Metr. 1: 4.

4. Yabroff KR, Lamont EB, Mariotto A, et al. (2008) Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst. 100: 630-41.

5. Verdecchia A, Capocaccia R, Egidi V, Golini A. (1989) A method for the estimation of chronic disease morbidity and trends from mortality data. Stat Med. 8: 201-16.

6. Keiding N. (1991) Age-specific incidence and prevalence: a statistical perspective (with discussion). J R Statist Soc A. 154: 371-412.

7. Gail MH, Kessler L, Midthune D, Scoppa S. (1999) Two approaches for estimating disease prevalence from population-based registries of incidence and total mortality. Biometrics. 55: 1137-44.

8. Tracey E, Alam N, Chen W, Bishop J. (2008) Cancer in New South Wales: Incidence and Mortality 2006. Sydney: Cancer Institute NSW.

9. Jemal A, Siegel R, Xu J, Ward E. (2010) Cancer statistics, 2010. CA Cancer J Clin. 60: 277-300.

10.   De Angelis G, De Angelis R, Frova L, Verdecchia A. (1994) MIAMOD: a computer package to estimate chronic disease morbidity using mortality and survival data. Comput Methods Programs Biomed. 44: 99-107.

11.   Grande E, Inghelmann R, Francisci S, et al. (2006) Estimating regional cancer burden in countries with partial registration coverage: an application to all malignant neoplasms in Italy over the period 1970-2010. Eur J Cancer. 42: 3236-45.

12.   De Angelis R, Grande E, Inghelmann R, et al. (2007) Cancer prevalence estimates in Italy from 1970 to 2010. Tumori. 93: 392-7.

13.   Australian Institute of Health and Welfare, Cancer Australia, Australasian Association of Cancer Registries. (2008) Cancer Survival and Prevalence in Australia: Cancers diagnosed from 1982 to 2004. Canberra: AIHW.

14.   Perme MP, Stare J, Esteve J. (2012) On estimation in relative survival. Biometrics. 68: 113-20.

15.   Brenner H, Gefeller O, Hakulinen T. (2004) Period analysis for 'up-to-date' cancer survival data: theory, empirical evaluation, computational realisation and applications. Eur J Cancer. 40: 326-35.

16.   De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. (1999) Mixture models for cancer survival analysis: application to population- based data with covariates. Stat Med. 18: 441-54.

17.   Yu B, Tiwari RC, Cronin KA, McDonald C, Feuer EJ. (2005) CANSURV: A Windows program for population-based cancer survival analysis. Comput Methods Programs Biomed. 80: 195-203.

18.   Lambert PC, Thompson JR, Weston CL, Dickman PW. (2007) Estimating and modeling the cure fraction in population-based cancer survival analysis. Biostatistics. 8: 576-94.

19.   Brenner H, Hakulinen T. (2004) Are patients diagnosed with breast cancer before age 50 years ever cured? J Clin Oncol. 22: 432-8.

Table 1: Study characteristics of the distant metastatic breast cancer cases and the breast cancer deaths, New South Wales females, 1980-2004

			Age group (years)							
			18-49		50-69		70-84		18-84	
Outcome	Category	Level	n	(%)	n	(%)	n	(%)	n	(%)
Breast cancer deaths	Total	Total	3337	(100.0)	8615	(100.0)	6261	(100.0)	18413	(100.0)
	Year of death	1980-1984	549	(16.5)	1552	(18.0)	917	(14.6)	3018	(16.6)
		1985-1989	682	(20.4)	1715	(19.9)	1201	(19.2)	3598	(19.8)
		1990-1994	777	(23.3)	1815	(21.1)	1304	(20.8)	3896	(21.4)
		1995-1999	708	(21.2)	1740	(20.2)	1403	(22.4)	3851	(21.1)
		2000-2004	621	(18.6)	1793	(20.8)	1436	(22.9)	3850	(21.1)
	Stage at diagnosis	Localised	877	(30.3)	2419	(32.6)	1919	(37.8)	5215	(33.9)
		Regional	1687	(58.2)	3953	(53.3)	2280	(44.9)	7920	(51.5)
		Distant	334	(11.5)	1038	(14.0)	881	(17.3)	2253	(14.6)
		Unknown	439		1205		1181		2825	
Metastatic diagnoses	Total	Total	3943	(100.0)	6173	(100.0)	2455	(100.0)	12571	(100.0)
	Year of diagnosis	1980-1984	895	(22.7)	1636	(26.5)	649	(26.4)	3180	(25.3)

28

			Col1	(%)	Col2	(%)	Col3	(%)	Col4	(%)
		1985-1989	297	(7.5)	532	(8.6)	262	(10.7)	1091	(8.7)
		1990-1994	332	(8.4)	546	(8.8)	251	(10.2)	1129	(9.0)
		1995-1999	942	(23.9)	1288	(20.9)	479	(19.5)	2709	(21.5)
		2000-2004	1477	(37.5)	2171	(35.2)	814	(33.2)	4462	(35.5)
	Stage at initial diagnosis	Localised	1107	(31.8)	1655	(30.2)	542	(24.9)	3304	(29.7)
		Regional	1724	(49.6)	2357	(43.0)	640	(29.4)	4721	(42.4)
		Distant	647	(18.6)	1470	(26.8)	994	(45.7)	3111	(27.9)
		Unknown	465		691		279		1435	
Breast cancer deaths with metastatic episodes	Total	Total	1896	(56.8)	5062	(58.8)	3189	(50.9)	10147	(55.1)
(% of breast cancer deaths)	Year of diagnosis	1980-1984	519	(94.5)	1454	(93.7)	815	(88.9)	2788	(92.4)
		1985-1989	283	(41.5)	681	(39.7)	399	(33.2)	1363	(37.9)
		1990-1994	158	(20.3)	458	(25.2)	298	(22.9)	914	(23.5)
		1995-1999	396	(55.9)	899	(51.7)	606	(43.2)	1901	(49.4)
		2000-2004	540	(87.0)	1570	(87.6)	1071	(74.6)	3181	(82.6)

Table 2: Relative survival estimates, metastatic breast cancer, NSW females 1980-2004

Age group (years)	Metastatic diagnosis category	Time (years)	Number at risk	Number of events	Survival	(95% CI)
18-49 years	Initial	0	647	0	1	(1.000, 1.000)
		1	466	148	0.768	(0.736, 0.802)
		2	326	93	0.608	(0.570, 0.648)
		5	138	114	0.367	(0.327, 0.413)
		10	57	40	0.238	(0.198, 0.287)
	Subsequent	0	2074	0	1	(1.000, 1.000)
		1	940	1035	0.498	(0.477, 0.521)
		2	551	322	0.321	(0.301, 0.343)
		5	171	267	0.147	(0.130, 0.165)
		10	76	47	0.101	(0.085, 0.118)
50-69	Initial	0	1470	0	1	(1.000, 1.000)
		1	900	509	0.654	(0.629, 0.679)
		2	647	197	0.51	(0.484, 0.538)
		5	224	245	0.285	(0.259, 0.314)
		10	90	66	0.191	(0.164, 0.223)
	Subsequent	0	4762	0	1	(1.000, 1.000)
		1	1980	2566	0.463	(0.449, 0.478)
		2	1192	644	0.3084	(0.295, 0.323)
		5	322	633	0.1259	(0.115, 0.138)
		10	122	120	0.0746	(0.065, 0.086)
70-84	Initial	0	994	0	1	(1.000, 1.000)
		1	419	519	0.489	(0.458, 0.523)

	2	258	128	0.348	(0.317, 0.383)
	5	66	124	0.168	(0.139, 0.204)
	10	11	27	0.095	(0.063, 0.143)
Subsequent	0	2624	0	1	(1.000, 1.000)
	1	819	1652	0.3789	(0.360, 0.399)
	2	483	248	0.2631	(0.245, 0.283)
	5	117	233	0.1211	(0.105, 0.139)
	10	18	46	0.0615	(0.045, 0.085)

Table 3: MIAMOD predictions for the incidence and prevalence of distant metastatic breast cancer, Australian women, 2004

Age group (years)	Population (n)	Predicted incidence (n)	Predicted mortality (n)	Predicted prevalence by years since metastasis detection (n)					
				Total	0-4	5-9	10-14	15-19	20+
18-29	1618222	12	7	16	15	1	0	0	0
30-34	765164	34	22	57	51	5	1	0	0
35-39	735150	82	56	156	133	18	4	1	0
40-44	774055	171	124	366	293	53	15	4	1
45-49	721236	256	198	621	461	104	38	13	5
50-54	664728	328	273	870	579	166	76	34	16
55-59	596345	363	309	1044	636	188	114	62	43
60-64	448058	317	275	1006	567	173	106	80	80
65-69	377972	304	272	1034	540	172	106	81	135
70-74	326994	307	278	1027	472	171	110	83	192
75-79	302107	353	302	1118	500	171	111	87	249
80-84	230308	375	310	969	450	138	86	67	228
18-84	7560339	2902	2426	8284	4696	1358	767	512	949

Table 4: MIAMOD predictions for the prevalence of distant metastatic breast cancer, Australian Women, 2004

Age group (years)	Ratio of prevalence to breast cancer deaths	Distribution of 100 prevalent distant metastatic cases by years since metastasis detection (for an age group)				
		0-4	5-9	10-14	15-19	20+
18-29	1.40	94	6	0	0	0
30-34	2.59	89	9	1	0	0
35-39	2.77	85	12	3	1	0
40-44	2.96	80	14	4	1	0
45-49	3.13	74	17	6	2	1
50-54	3.19	67	19	9	4	2
55-59	3.38	61	18	11	6	4
60-64	3.66	56	17	11	8	8
65-69	3.80	52	17	10	8	13
70-74	3.70	46	17	11	8	19
75-79	3.71	45	15	10	8	22
80-84	3.12	46	14	9	7	24

Fig. 1: Relative survival for women with metastatic disease, by age and initial or subsequent metastatic disease, NSW, 1980-2004

Fig. 2: Estimated numbers of metastatic breast cancer cases using the counting method (June 2004) and MIAMOD (2004), NSW females aged 18-84 years